



Client Overview

- The client is one of the top messaging service providers running multiple channels on major Social Media Messaging Platforms. Client was interested in creating an automated process to filter and classify messages into SPAM and HAM (legitimate messages that needs to be sent to a respective recipient) so that they can flag the SPAM messages and subject it to a more targeted verification.



Project Overview

- The project leveraged teX.ai to perform text analytics on data fed from messaging channels. The developed system automated the extraction of data from the channels, classification of the messages into HAM and SPAM, along with the alerting mechanisms for downstream actions.



Business Requirement

- Access the communication channels to read the messages being posted
- Filter the HAM messages from the SPAM messages
- Flag the SPAM messages for further verification
- Optimize the process and automate the system



Solution

- Data was labelled into SPAM and HAM, before being used for training the model, embracing supervised learning methods
- Random forest, Naive Bayes, and SVM algorithms were used and their outcomes optimized so the most suitable algorithm was selected for maximum accuracy
- A robust retraining process was deployed to enable the Machine Learning model to continuously relearn with each new data point, thereby improving its accuracy over time
- An alerting mechanism was setup to notify the stakeholders when a SPAM message is flagged
- The various pieces, comprising of the data extraction process, Machine Learning model, alerting mechanism etc. was integrated into a single system and automated
- Name entity recognition was used to highlight the keywords and phrases in every section of the message



Technology

- SVM, Random Forest, Naïve Bayes



Business Impact

- The Machine Learning assisted classification process enabled message filtering into HAM and SPAM with a very high degree of accuracy, improved message quality by over 80%
- The automated classification process reduced the time and effort required for monitoring the messaging channels by 75%
- The end-to-end system enable rapid scaling of the number of communication channels with the same existing resources
- Text analytics enabled nearly a 60% reduction in costs and improved mechanisms to optimize the resources being deployed



INDIA

Chennai | Bengaluru | Mumbai
Toll-free: 1800-123-1191

USA

Cupertino | Princeton | Boston
Toll-free: +1-888-207-5969

UK

London
+44 773 653 9098

SINGAPORE

+65 9630 7959

MALAYSIA

Kuala Lumpur
+60 (3) 2298 8465